Bitcoin (BTC) Price:
A Predictive Analysis

Springboard - DSC | Capstone Project 2

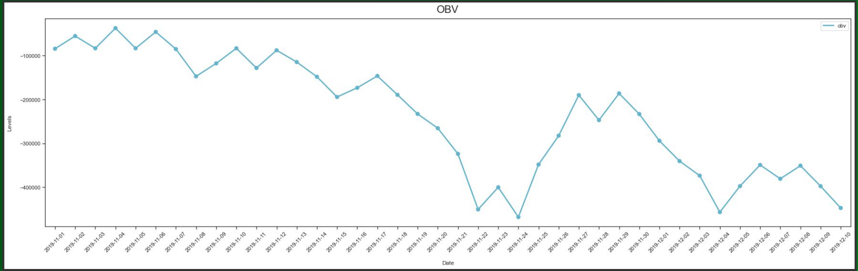
By John Arancio

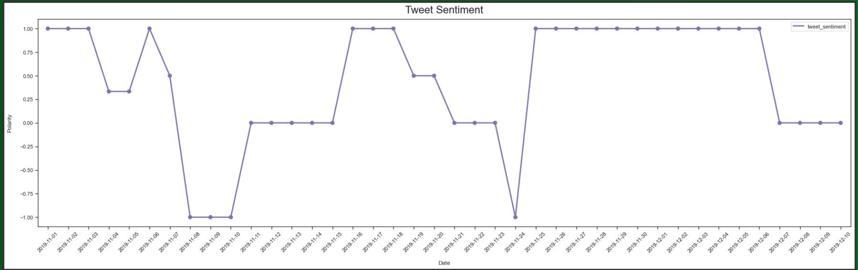
Can BTCs next day closing price be accurately predicted?

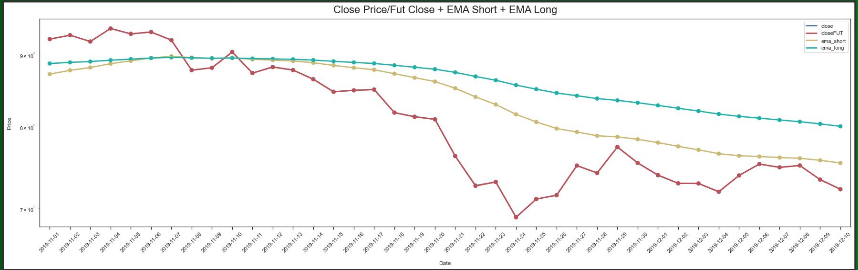
- Price action lacks understanding
  - Dramatic speculation
  - Lack of empirical reasoning
- Volatility
  - High risk
- Influx of institutional investment/interest
  - Grayscale Trust (owns 3.9% of circulating BTC)
  - Goldman Sachs (61% of clients expect to increase exposure to cryptocurrency)

Approach

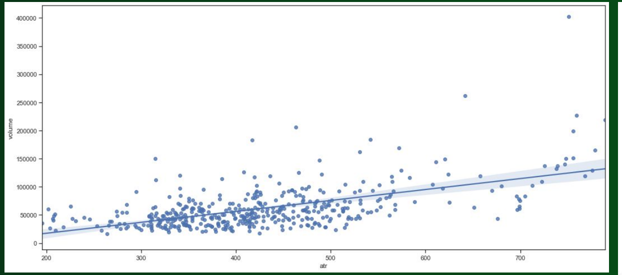
- Binance API client (historical daily prices) [09/02/2019 - 11/15/2020]
    - Range: $10,340 - $15,957 (+54%)
    - Low: $4,800
    - High: $16,320.70
    - Avg. Price: $9,367.31
- Indicators as potential features
    - EMA Long/Short, ATR, OBV
    - Total indicators analyzed: 6
- Tweet sentiment via historical web scrape
    - Avg. sentiment per day (mean polarity)
        - Bullish (positive) tweet = 1
        - Neutral tweet = 0
        - Bearish (negative) tweet = -1

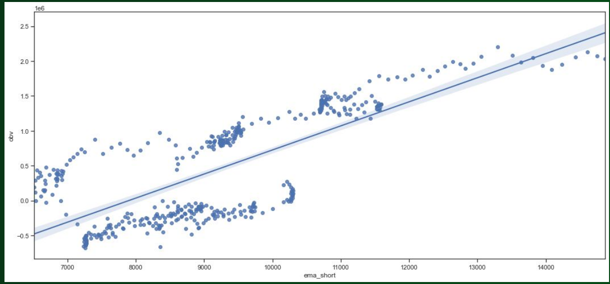


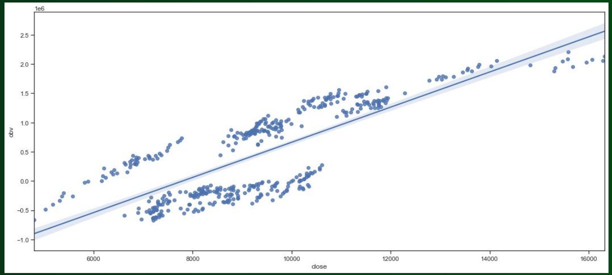


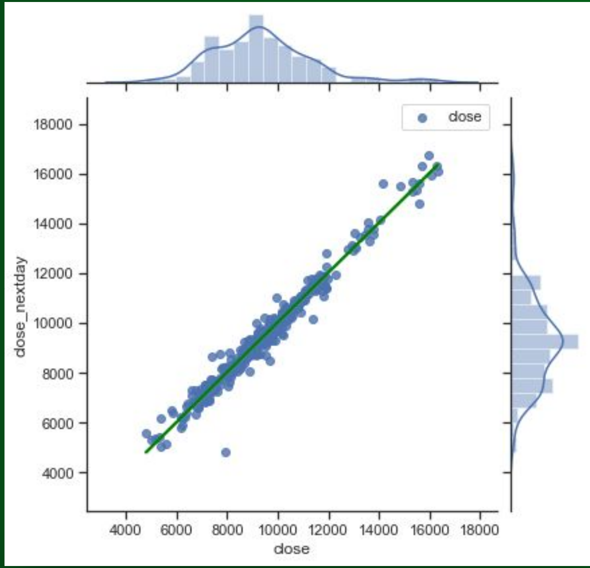


Approach









Exploratory Data Analysis

and

Feature Selection

- ฿ Variable multicollinearity
  - Volume/ATR
  - OBV/EMA Short
  - OBV/Close
- ฿ Correlation to target
  - Highest linear correlation
    - Close
    - EMA Long/Short
    - OBV
    - Volume
    - ATR
    - Tweet Sentiment

Modeling

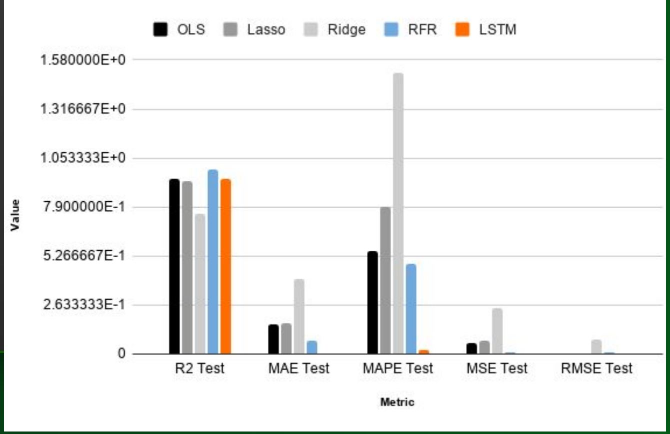
- Five distinct models
  - Ordinary Least Squares (OLS)
  - Lasso
  - Ridge
  - Random Forest Regression (RFR)
  - Long Short Term Memory (LSTM)
- Model Tuning
  - Regression Pipeline
    - GridsearchCV
  - RNN
    - Hyperas
- Interpretation
  - LIME Explainer
  - Residuals
  - Coefficients
  - Feature Importance

Linear Regression

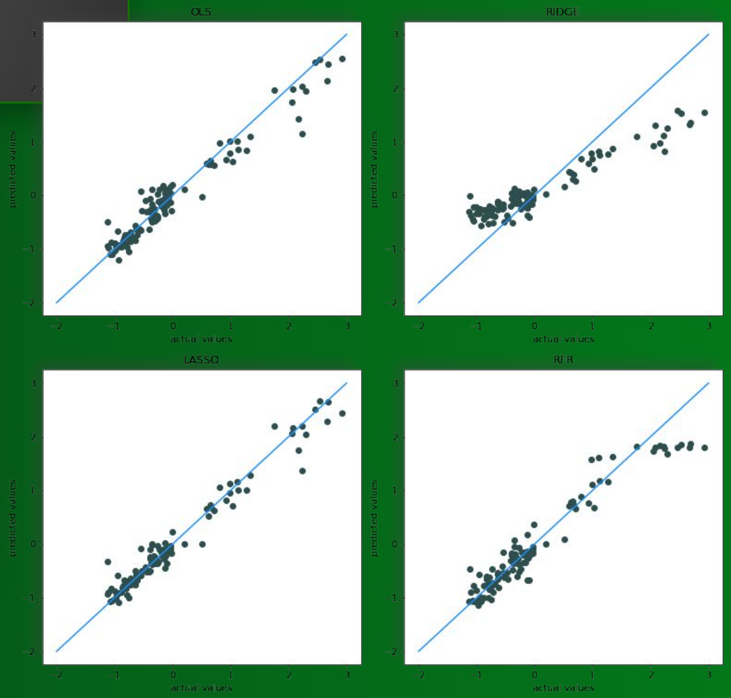
Ensemble Regression (Decision Tree)

Recurrent Neural Network (RNN)

- Best Performance
  - Random Forest Regression (RFR)
  - Long Short Term Memory (LSTM)
- Worst Performance
  - Ridge

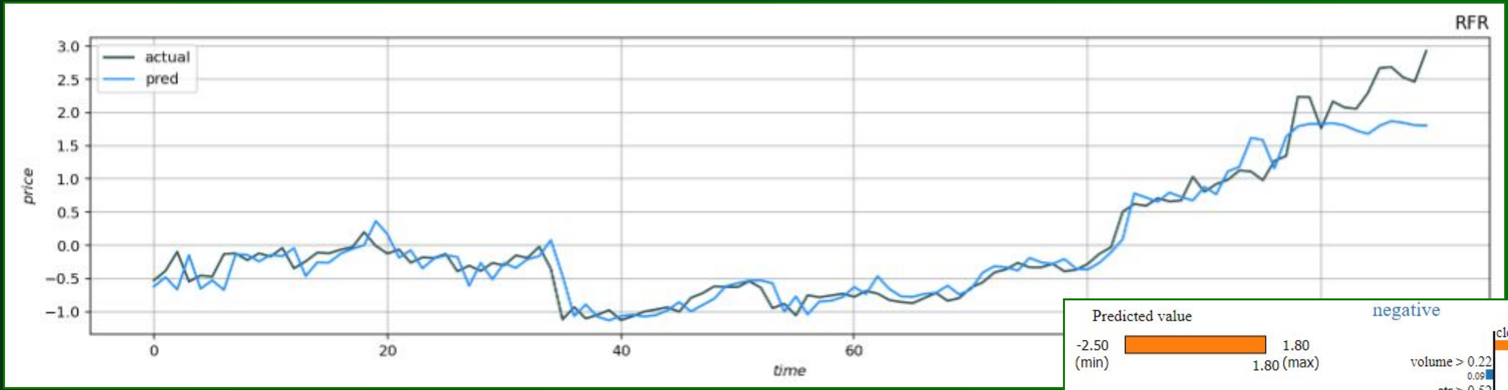


Regression Models Performance: Actual Vs Pred

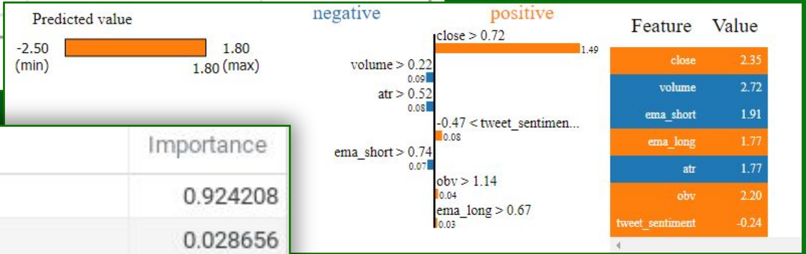


Findings

(RFR Test Prediction)



	R2 Test	MAE Test	MAPE Test	MSE Test	RMSE Test	OOB Train
RFR	0.98845	0.06853	0.48569	0.01176	0.00686	0.91361



		Importance
1	close	0.924208
2	atr	0.028656
3	obv	0.013587
4	ema_long	0.013533
5	volume	0.011165
6	ema_short	0.00675

- Scores
  - Highest R2
  - Lowest MAE, MSE, RMSE
  - Valid OOB
- Interpretation
  - Close HEAVILY weighted
  - Other features minimally weighted

Findings

(LSTM Test Prediction)

	R2 Test ▲	MAE Test	MAPE Test	MSE Test	RMSE Test
LSTM	0.9395	304.9840	0.0247	180,906.31	425.3308

- ฿ Scores
  - Second highest R2
  - Lowest MAPE
  - MAE = approx. $305 avg difference
- ฿ Residuals
  - Greatest positive: 1,129.17
  - Greatest negative: -1,730.21

Conclusions

- ฿ Volatility handled well by RFR and LSTM
  - Volume momentum plays big role in BTC price action (OBV)
  - ATR useful with RFR
- ฿ LSTM creates dynamic S&R (support and resistance) line, similar to a EMA/MA
- ฿ Models generate confluence for potential signals
  - Risk reduction
- ฿ Standard asset indicators/metrics (i.e. open, close, volume) more important than extra indicators

Further Work

- ➢ Perform deeper analysis into Tweet sentiment
- ➢ Train/test models on more data
- ➢ Fine tune the hyperparameters of models
- ➢ Perform EDA with other indicators
- ➢ Implement signal generator
  - Back test
  - Compute Sharpe Ratio
- ➢ Test for replicable results on different assets (equities, ETFs, etc.)

Recommendations

- Use RFR and LSTM greatest positive and negative residual values to adjust model
- Use top performing models as a signal generator in conjunction with own due diligence
- Implement automated strategy to run LSTM as its own “advisor” or “HF trader”
- Use findings to improve previous analysis or speculation for identifying potential volatile spikes